Propensity Scores

Todd Wagner, PhD

October 2013

Health Economics Resource Center

HSR&D
HEALTH SERVICES RESEARCH & DEVELOPMENT

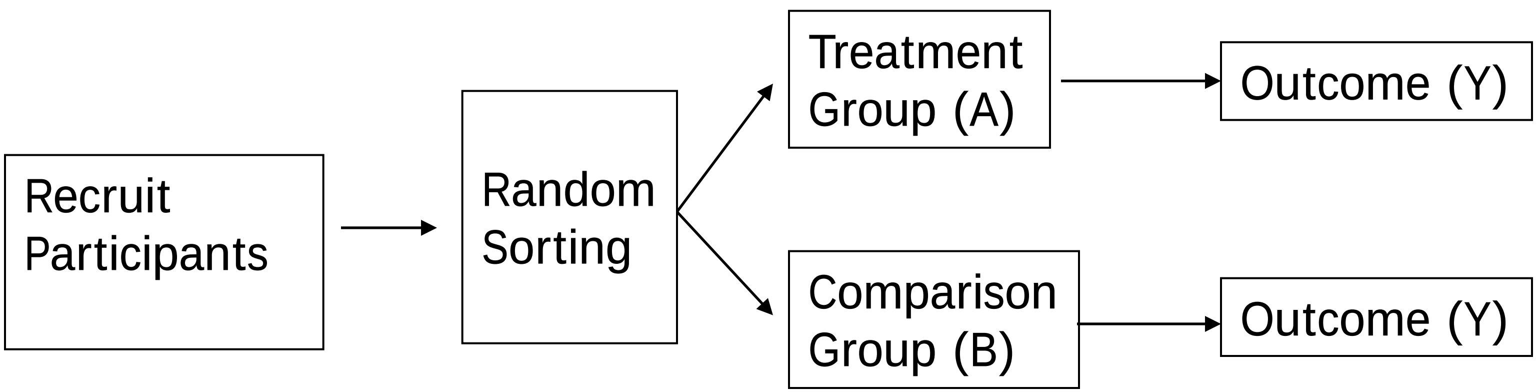
Outline

1. Background on assessing causation
2. Define propensity score (PS)
3. Calculate the PS
4. Use the PS
5. Limitations of the PS

Causality

- Researchers are often interested in understanding causal relationships
  - Does drinking red wine affect health?
  - Does a new treatment improve mortality?
- Randomized trial provides a methodological approach for understanding causation

Randomization



Note: random sorting can, by chance, lead to unbalanced groups. Most trials use checks and balances to preserve randomization

Trial analysis

- The expected effect of treatment is

  $E(Y)=E(Y^A)-E(Y^B)$

  Expected effect on group A minus expected effect on group B (i.e., mean difference).

Trial Analysis (II)

- $E(Y)=E(Y^A)-E(Y^B)$ can be analyzed using the following general model

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Where

- y is the outcome
- α is the intercept
- x is the mean difference in the outcome between treatment A relative to treatment B
- ε is the error term
- i denotes the unit of analysis (person)

Trial Analysis (III)

- The model can be expanded to control for baseline characteristics

$$y_i = \alpha + \beta x_i + \delta Z_i + \varepsilon_i$$

Where

- y is outcome
- α is the intercept
- x is the added value of the treatment A relative to treatment B
- Z is a vector of baseline characteristics (predetermined prior to randomization)
- ε is the error term
- i denotes the unit of analysis (person)

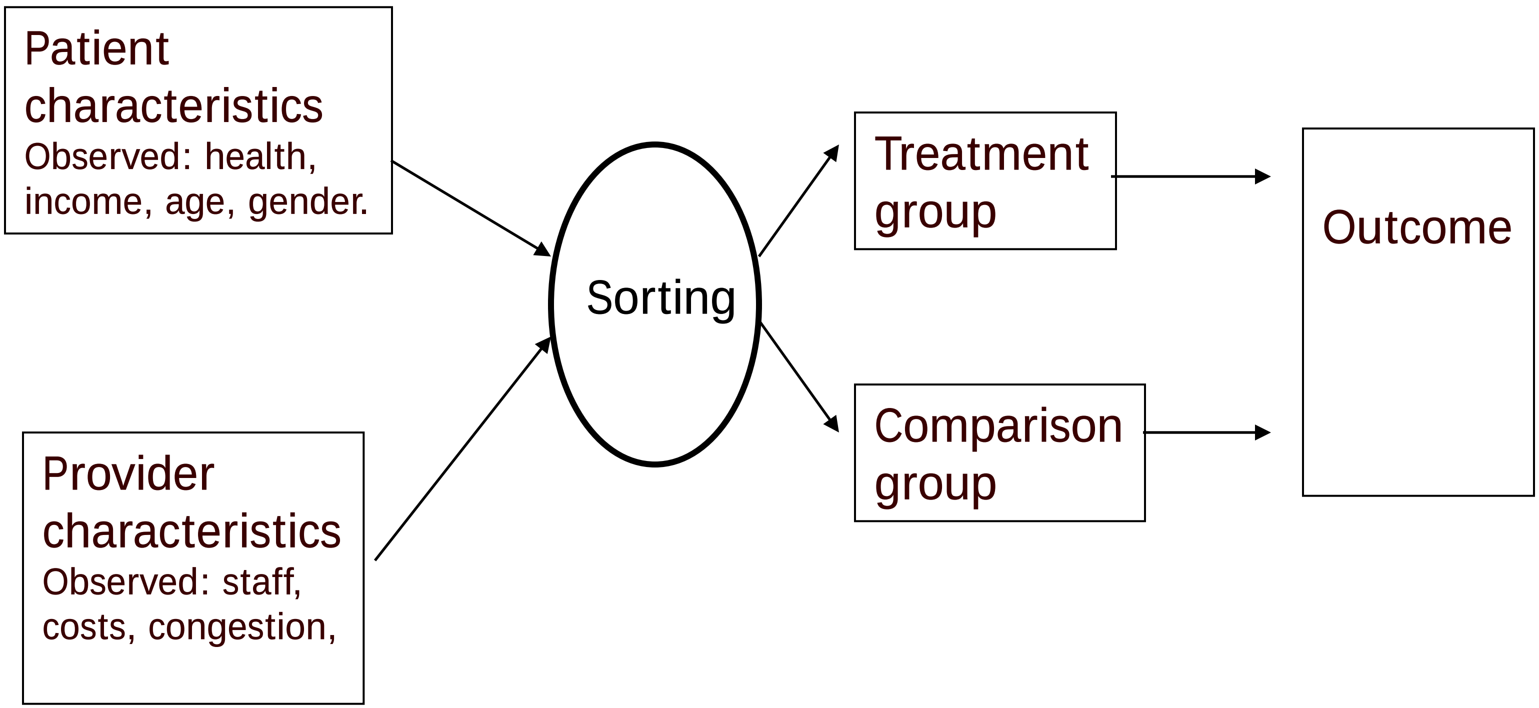
Assumptions

- Right hand side variables are measured without noise (i.e., considered fixed in repeated samples)
- There is no correlation between the right hand side variables and the error term $E(x_i u_i)=0$
- If these conditions hold, β is an unbiased estimate of the causal effect of the treatment on the outcome

Observational Studies

- Randomized trials may be
  - Unethical
  - Infeasible
  - Impractical
  - Not scientifically justified

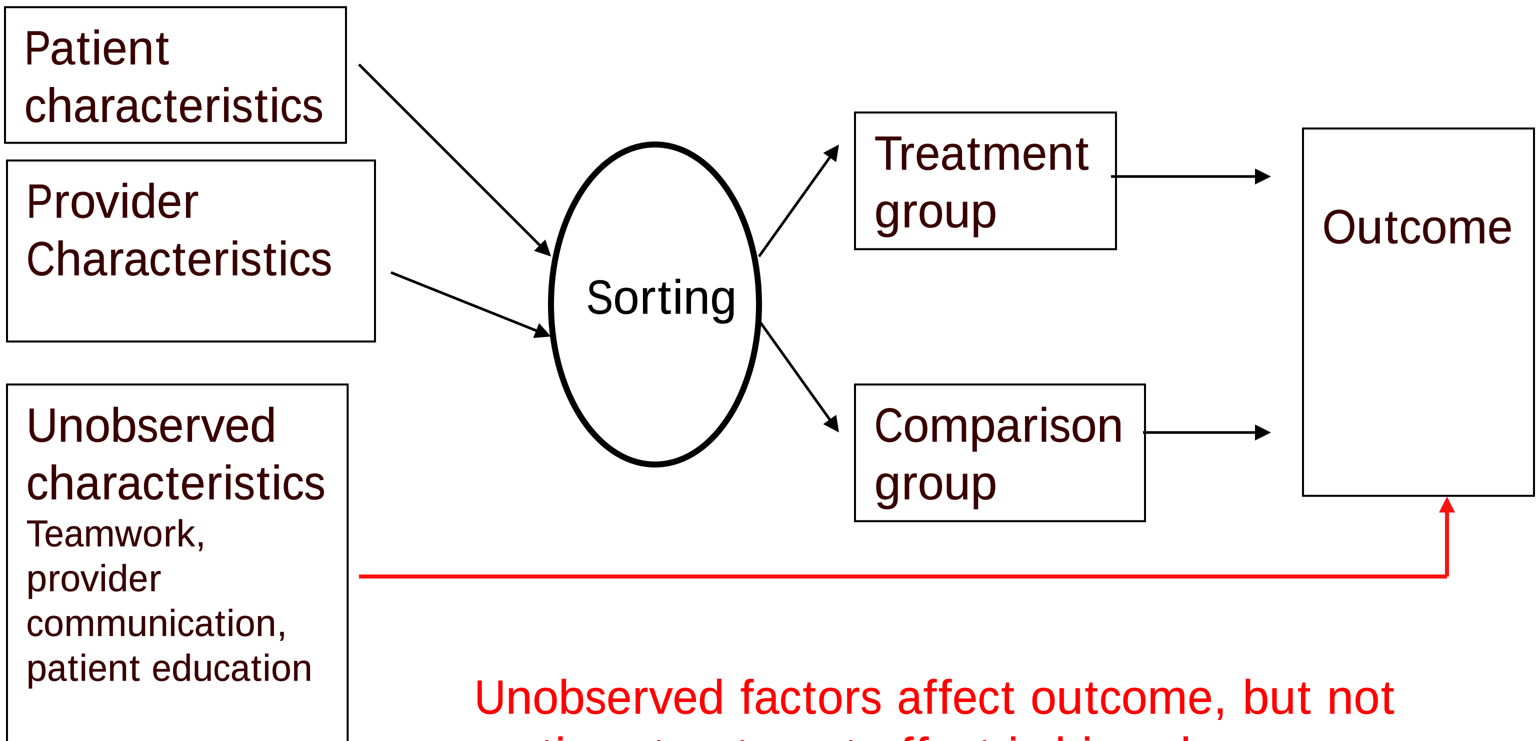
Sorting without randomization



If everything is fully observed and correctly specified; results are not biased. Never happens in reality.

Based on: Maciejewski and Pizer (2007) Propensity Scores and Selection Bias in Observational Studies. HERC Cyberseminar

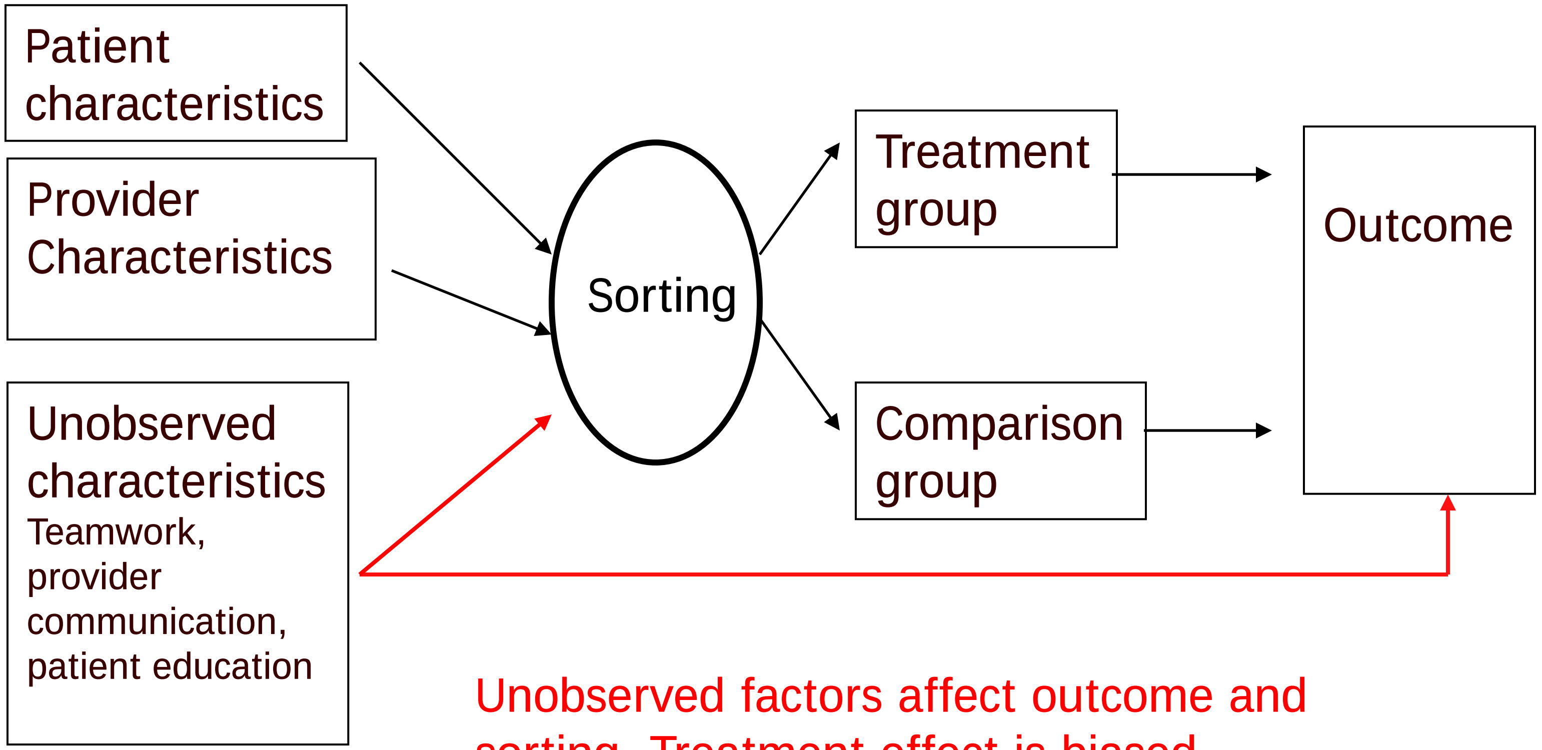
Sorting without randomization



Unobserved factors affect outcome, but not sorting; treatment effect is biased.

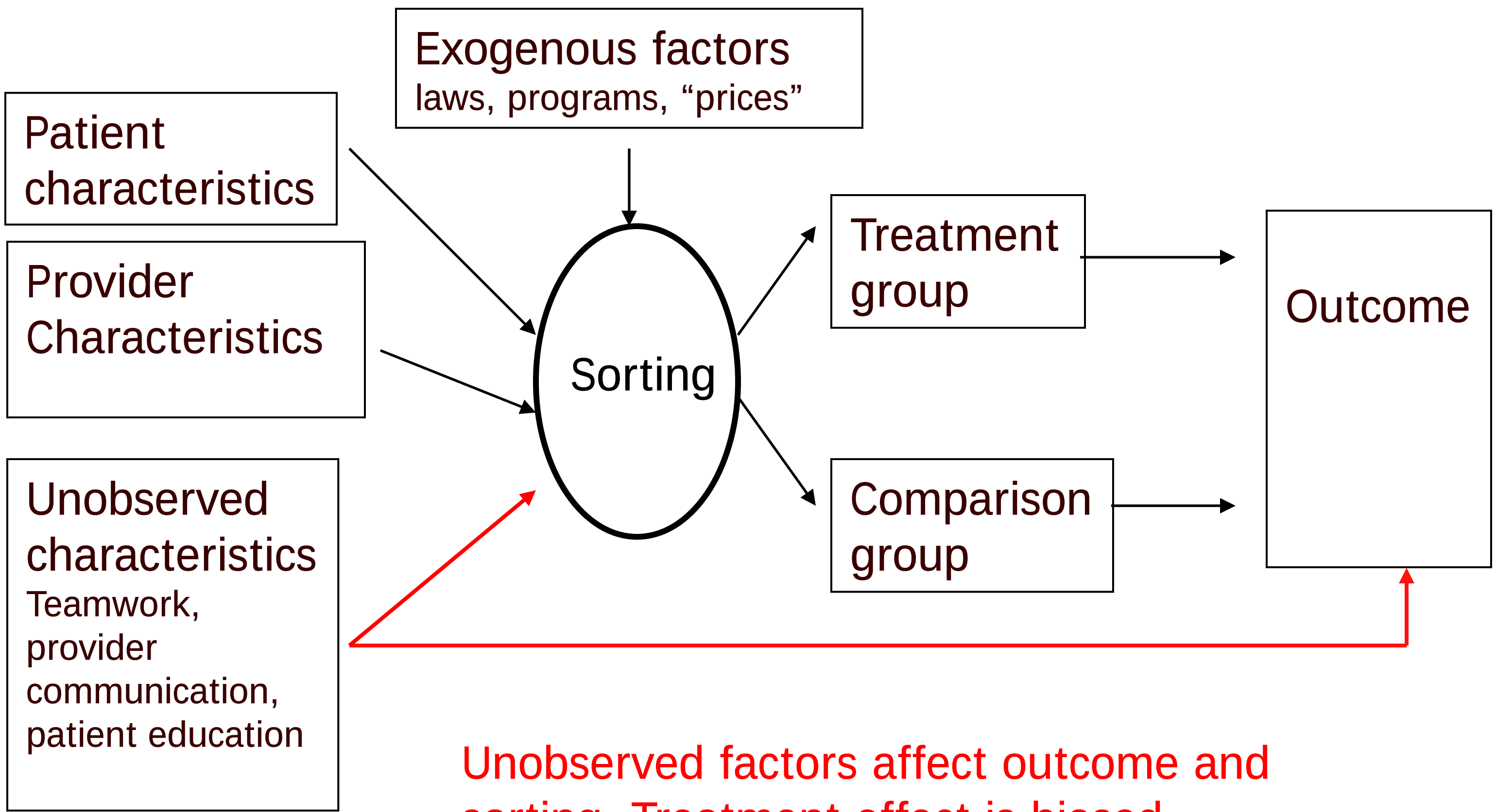
Fixed effects would be potential fix.

Sorting without randomization



Unobserved factors affect outcome and sorting. Treatment effect is biased. Causality isn't identified.

Sorting without randomization



Unobserved factors affect outcome and sorting. Treatment effect is biased.

Instrumental variables may offer insights on causal relationship, as related to exogenous factors.

Propensity Score Defined

- The PS uses observed information, which is multi-dimensional, to calculate a single variable (the score)
- The score is the predicted propensity to get sorted (usually thought of as propensity to get treatment).

  Expected treatment effect: $E(Y)=E(Y^A)-E(Y^B)$

  Propensity Score is: $Pr(Y=A \mid X_i)$

Propensity Scores

- What it is: Another way to correct for observable characteristics

- What it is not: A way to adjust for unobserved characteristics

- The only way to make causal claims is to make **huge** assumptions.

Strong Ignorability

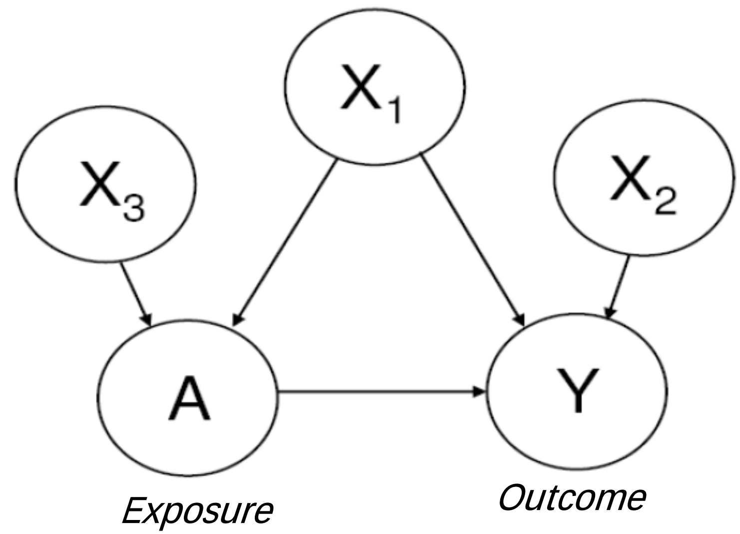
- To make statements about causation, you would need to make an assumption that treatment assignment is strongly ignorable.
  - Similar to assumptions of missing at random
  - Equivalent to stating that all variables of interest are observed

Calculating the Propensity Score

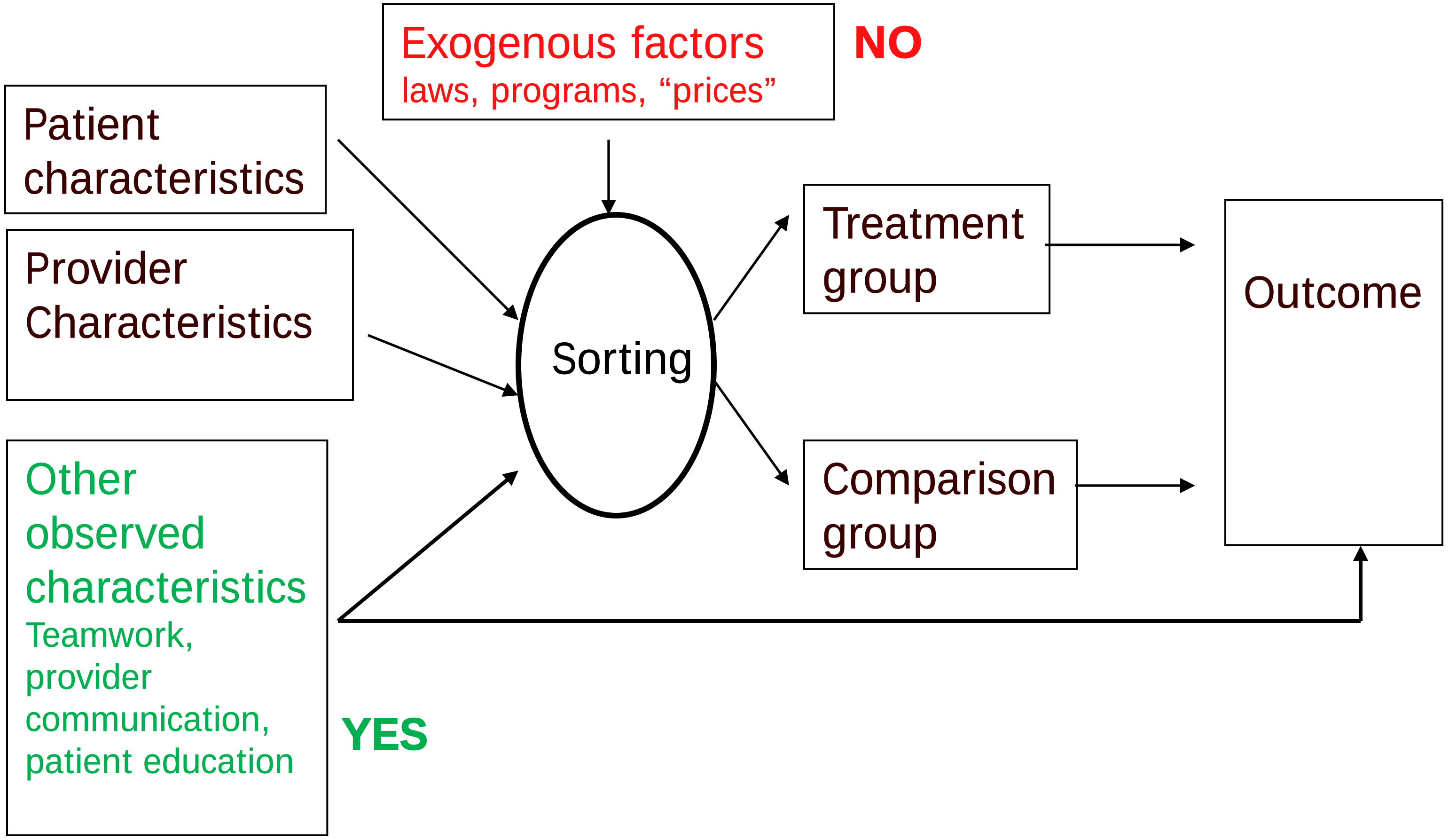
- One group receives treatment and another group doesn’t
- Use logistic regression to estimate the probability that a person received treatment
- The predicted probability is the propensity score

Variables to Include

- Include variables that are unrelated to the exposure but related to the outcome
- This will decrease the variance of an estimated exposure effect without increasing bias



Variables to Include in PS



Variables to Exclude

- Exclude variables that are related to the exposure but not to the outcome
- These variables will increase the variance of the estimated exposure effect without decreasing bias
- Variable selection is particularly important in small studies (n<500)

Brookhart MA, et al Am J Epidemiol. 2006 Jun 15;163(12):1149-56. Variable selection for propensity score models.

How do You Use a Propensity Score

Uses

- Understanding sorting and balance
  - Sorting is multidimensional
  - The PS provides a simple way of reducing this dimensionality to understand the similarity of the treatment groups
- Adjusting for covariance

Example

- Are surgical outcomes worse when the surgeon is a resident?
- Resident assignment may depend on
  - Patient risk
  - Availability of resident
  - Resident skill
  - Local culture

Resident Assignment

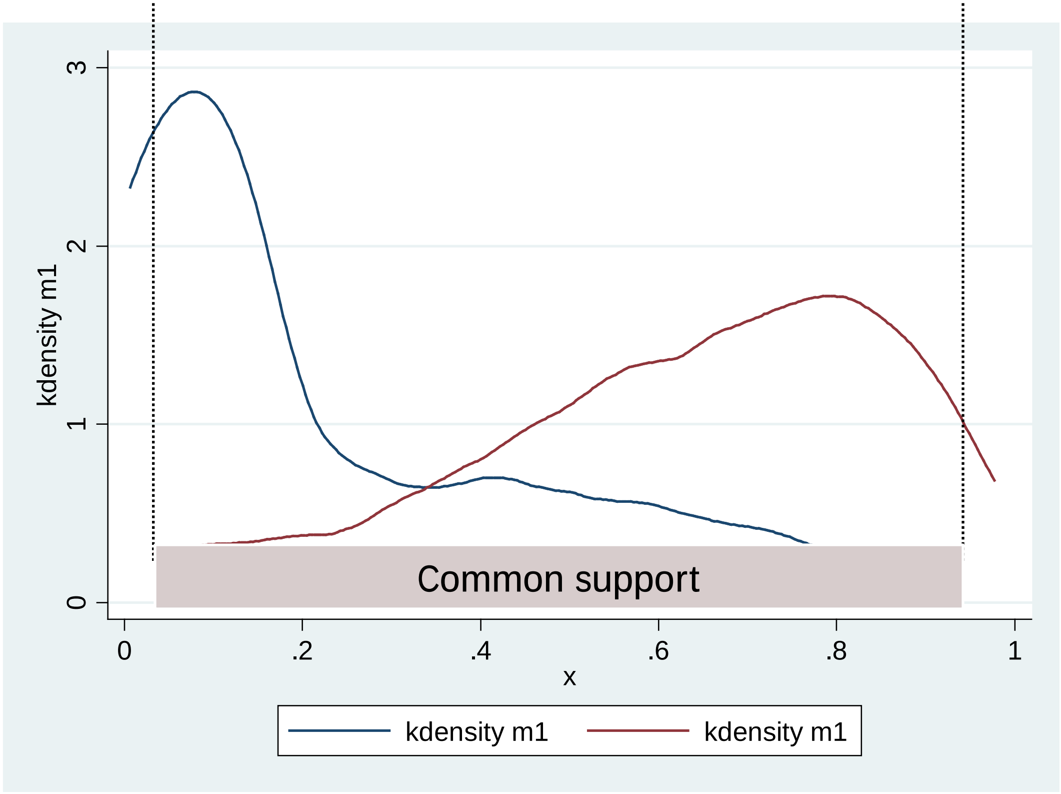
	OR	P value
Age	1.00	0.79
Canadian Functional Class		
Class 2	1.93	0.15
Class 3	2.12	0.09
Class 4	4.25	0.02
Urgent priority	0.93	0.89
Artery condition at site		
Calcified	0.67	0.25
Sclerotic	2.63	0.00
site 2	62.89	0.00
site 3	0.67	0.60
site 5	138.16	0.00
site 7	11.66	0.00
site 8	19.85	0.00
site 9	1.76	0.43
endo vascular harvest	0.20	0.01
On pump surgery	1.20	0.75
1-2 grafts	1.70	0.16
4-5 grafts	0.79	0.46

Assignment not associated with age or number of grafts

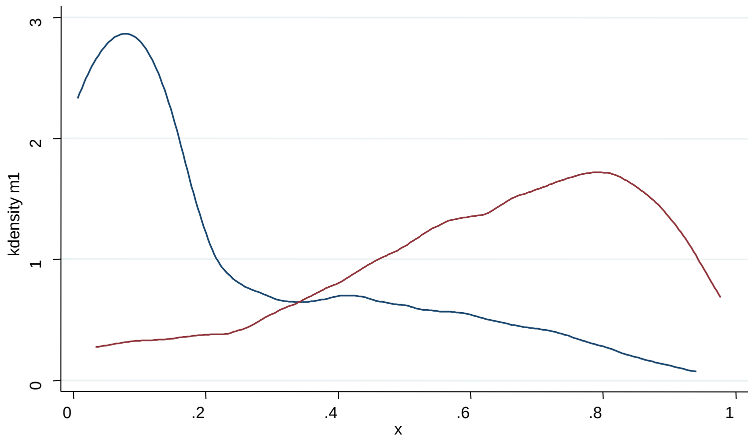
Assignment associated with angina symptoms and planned harvesting technique

Bakaeen F, Sethi G, Wagner T, et al. Coronary Artery Bypass Graft Patency: Residents Versus Attending Surgeons. *Annals of Thoracic Surgery.*

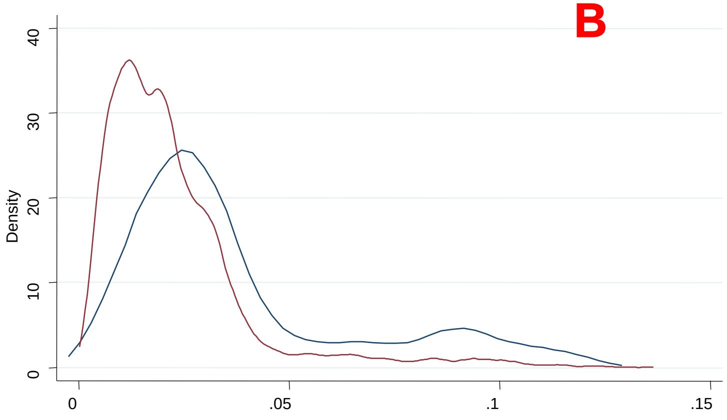
Propensity Score for Resident vs Attending Surgeon



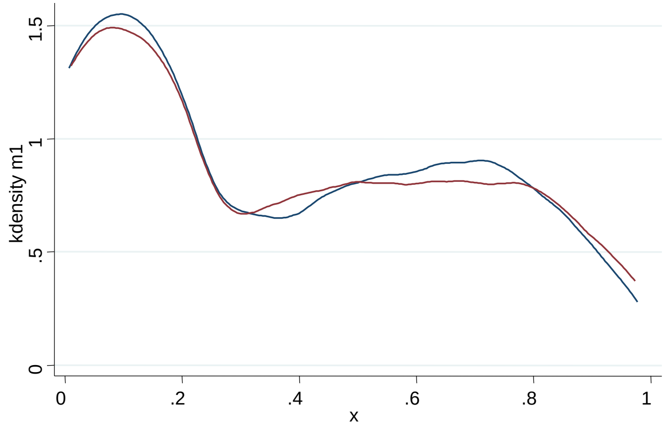
Compare Three Scores



B



C



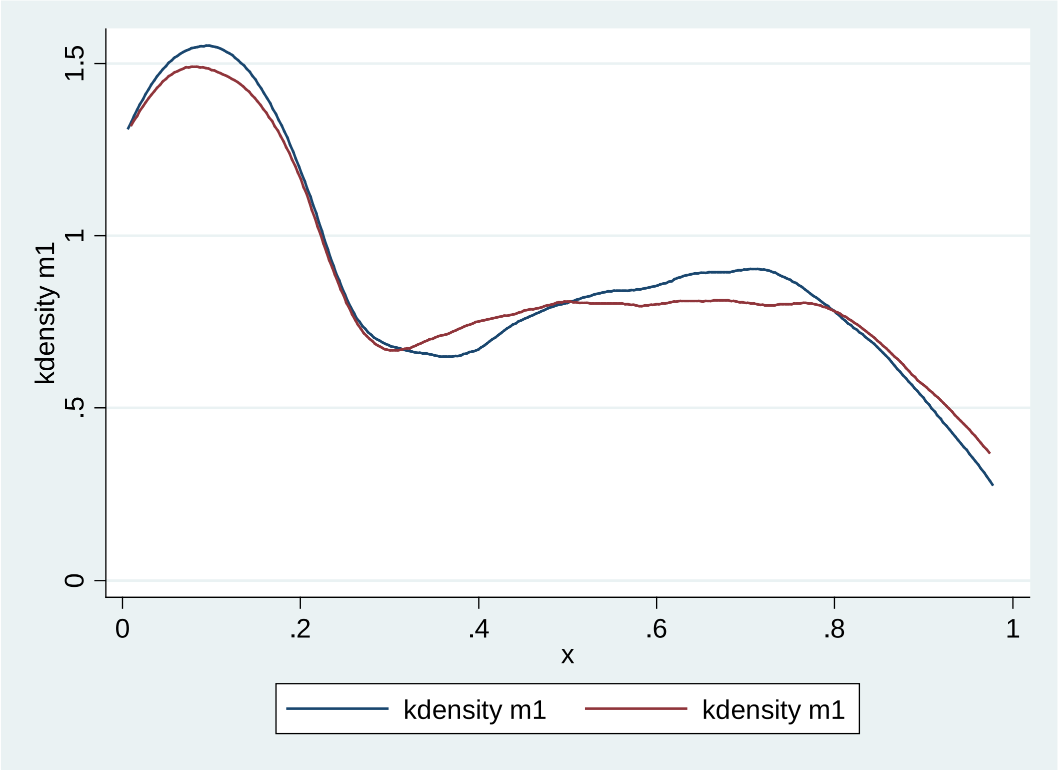
Poll

- Do any of these distributions concern you? Choose one
- A
- B
- C
- All of them
- None of them

RCTs and Propensity Scores

- What would happen if you used a propensity score with data from a RCT?

Share Common Support



Using the Propensity Score

1. Compare individuals based on similar PS scores (a matched analysis)
2. Conduct subgroup analyses on similar groups (stratification)
3. Include it as a covariate (quintiles of the PS) in the regression model
4. Use it to weight the regression (i.e., place more weight on similar cases)
5. Use both 3 and 4 together (doubly robust)

Matched Analyses

- The idea is to select controls that resemble the treatment group in all dimensions, except for treatment
- You can exclude cases and controls that don't match, which can reduce the sample size/power.
- Different matching methods

Matching Methods

- Nearest Neighbor: rank the propensity score and choose control that is closest to case.
- Caliper: choose your common support and from within randomly draw controls

PS as a Covariate

- There seems to be little advantage to using PS over multivariate analyses in most cases.[1]
- PS provides flexibility in the functional form
- Propensity scores may be preferable if the sample size is small and the outcome of interest is rare.[2]

1. Winkelmeyer. Nephrol. Dial. Transplant 2004; 19(7): 1671-1673.
2. Cepeda et al. Am J Epidemiol 2003; 158: 280–287

Doubly Robust Estimators

Expected treatment effect: $E(Y)=E(Y^A)-E(Y^B)$

1. Fit a logistic regression model for treatment conditional on the baseline variables. The predicted value from this regression gives the estimated propensity scores (PS_i)

2. Fit a regression model for outcome (Y_i) on the baseline variables for the treatment group only ($Y_i = A$), and obtain the predicted values for the whole sample.

3. Fit the same regression model for outcome on the baseline variables for the control group only ($Y_i = B$), and obtain the predicted values for the whole sample.

4. Plug the PS_i, Pred (A), and Pred(B) into a formula for the double-robust estimator (essentially a PS weighted mean difference) and bootstrap the SE.

Emsley R, Lunt M, Pickles A, Dunn G Implementing double-robust estimators of causal effects The Stata Journal (2008) 8, Number 3, pp. 334–353, http://www.stata-journal.com/sjpdf.html?articlenum=st0149

Doubly Robust Estimators

- Have gained favor because DR provides some protection from mis-specification in either the regression or PS model.

Tsiatis, A. A. 2006. *Semiparametric Theory and Missing Data. New York: Springer.*
Leon, S., A. A. Tsiatis, and M. Davidian. 2003. Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics 59: 1046–1055.*

Limitations

Do the Unobservables Matter?

- Propensity scores focus only on observed characteristics, <u>not on unobserved</u>.
- Improbable that we fully observe the sorting process
  - Thus $E(x_i u_i) \neq 0$
  - Multivariate (including propensity score) is biased and we need instrumental variables, fixed effects or RCT

Does Using PS Exacerbate Imbalance of Unobservables

- PS is based on observables.

- Brooks and Ohsfeldt, using simulated data, showed that PS models can create **greater** imbalance among unobserved variables.

Brooks and Ohsfeldt (2013): Squeezing the balloon: propensity scores and unmeasured covariate balance. HSR.

Summary

Overview

- Propensity scores offer another way to adjust for confound by observables
- Reducing the multidimensional nature of confounding can be helpful
- There are many ways to implement propensity scores and a growing interest in doubly robust estimators

Strengths

- Allow one to check for balance between control and treatment

- Without balance, average treatment effects can be very sensitive to the choice of the estimators.[1]

1. Imbens and Wooldridge 2007 http://www.nber.org/WNE/lect_1_match_fig.pdf

Challenges

- Propensity scores are often misunderstood
- Not enough attention is placed on the PS model, itself
- Not enough attention is placed on robustness checks
- While a PS can help create balance on observables, PS models do not control for unobservables or selection bias

Further Reading

- Imbens and Wooldridge (2007) ww.nber.org/WNE/lect_1_match_fig.pdf
- Guo and Fraser (2010) Propensity Score Analysis. Sage.
- Tsiatis, A. A. 2006. *Semiparametric Theory and Missing Data*. *New York: Springer.*
- Leon, S., A. A. Tsiatis, and M. Davidian. 2003. Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics 59: 1046–1055.*
- Rosenbaum, P. R., D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika 70: 41–55*
- Brookhart MA, et al Am J Epidemiol. 2006 Jun 15;163(12):1149-56. Variable selection for propensity score models.
- Brooks and Ohsfeldt (2013): Squeezing the balloon: propensity scores and unmeasured covariate balance. HSR.
- Emsley R, Lunt M, Pickles A, Dunn G Implementing double-robust estimators of causal effects The Stata Journal (2008) 8, Number 3, pp. 334–353, http://www.stata-journal.com/sjpdf.html?articlenum=st0149